Can appliances understand the behavior of elderly via machine learning? A feasibility study

Kun Qian, Tomoya Koike, Kazuhiro Yoshiuchi, Björn Schuller, Yoshiharu Yamamoto

Angaben zur Veröffentlichung / Publication details:

Qian, Kun, Tomoya Koike, Kazuhiro Yoshiuchi, Björn Schuller, and Yoshiharu Yamamoto. 2021. "Can appliances understand the behavior of elderly via machine learning? A feasibility study." *IEEE Internet of Things Journal* 8 (10): 8343–55. https://doi.org/10.1109/jiot.2020.3045009.

Can Appliances Understand the Behavior of Elderly Via Machine Learning? A Feasibility Study

Kun Qian, *Senior Member, IEEE*, Tomoya Koike, *Student Member, IEEE*, Kazuhiro Yoshiuchi, Björn W. Schuller, *Fellow, IEEE*, and Yoshiharu Yamamoto, *Member, IEEE*

***Abstract*—Over the last half decade, fast development of the Internet of Things and machine learning (ML) made it feasible to leverage the power of artificial intelligence to facilitate a variety of intelligent systems in smart home. Nevertheless, the studies on designing specific computing technologies for helping elderly to enjoy a comfortable, convenient, and independent daily life are extremely limited. On the one hand, there are increasingly growing demands from the ageing society to implement the cutting edge technology enabling a better life quality for the elderly. On the other hand, there is still a lack on fundamental investigations, applicable infrastructures, and advanced data-driven frameworks. To this end, we propose a novel machine framework for analyzing the daily life behavior of elderly—all in this study are living alone—by the data collected from their home appliances, i.e., television and refrigerator. First, the interevent intervals for the use of the appliances collected in one month from 76 elderly are the raw data to describe the behaviors. Then, three ML paradigms are investigated and compared, which include "classic" ML methods and the state-of-the-art deep learning approaches. Finally, we indicate the current findings and limitations in this feasibility study. Experimental results demonstrate that, our proposed method can reach performance peak at an unweighted average recall of 58.7 % (chance level: 50.0 %) in a subject-independent test for classifying symptom/nonsymptom days.**

***Index Terms*—Ageing society, context awareness, Internet of Things (IoT), machine learning (ML), smart appliances.**

This work was supported in part by the Zhejiang Lab's International Talent Fund for Young Professionals, China; in part by the Japan Society for the Promotion of Science Postdoctoral Fellowship for Research in Japan under Grant P19081; in part by the Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan, under Grant 19F19081, Grant 20H00569, and Grant 17H00878; and in part by the EU Horizon 2020 Research & Innovation Action Project under Grant 826506 (sustAGE). *(Corresponding author: Kun Qian.)*

Kun Qian, Tomoya Koike, and Yoshiharu Yamamoto are with the Educational Physiology Laboratory, Graduate School of Education, University of Tokyo, Tokyo 113-0033, Japan (e-mail: qian@p.u-tokyo.ac.jp; tommy@p.u-tokyo.ac.jp; yamamoto@p.u-tokyo.ac.jp).

Kazuhiro Yoshiuchi is with the Department of Stress Sciences and Psychosomatic Medicine, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan (e-mail: kyoshiuc-tky@umin.ac.jp).

Björn W. Schuller is with the Group on Language, Audio, and Music, Imperial College London, London SW7 2BU, U.K., and also with the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, 86159 Augsburg, Germany (e-mail: schuller@ieee.org).

I. Introduction

In THE beginning of this century, Schmidt and Laerhoven characterized *smart appliances* as the devices that are attentive to their environment [1]. In traditional fields, e.g., power consumption, smart appliances can contribute to the analysis of the consumers' demand responses via the help of a series of mathematical models [2]–[6]. On the one hand, within the fast development of the Internet of Things (IoT) and artificial intelligence (AI) over the last five years [7], tremendous work has been successfully applied in the fields, including *industrial environment surveillance* [8], *home management* [9], *smart buildings* [10], [11], *smart city* [12], *smart campus* [13], and *smart agriculture* [14]. These encouraging achievements make us confident that an era of AIoT (AI plus IoT) is coming.

On the other hand, the efforts leveraging the power of AI and IoT toward a ubiquitous and pervasive computing based on smart appliances for the personalized healthcare are still under way. In particular, the demand from the group of elderly has been increasingly raised since the ageing of the population is evident in all developed and many developing countries [15], [16]. Taking Japan (the world's oldest country [16]) as an example, approximately 27.6 % of the citizens are already 65 years old or even elder [17]. As a recent review literature indicated, AI has been found showing promising potential for ageing and longevity research in terms of biomarker discovery, personalized medicine, target identification, drug discovery, regenerative medicine, gene therapy, immuno oncology and immuno senescence, and many others [18].

Nevertheless, the choice of applicable smart appliances designed for bettering the life of the elderly is quite limited. To this end, we propose a novel machine learning (ML) framework for analyzing the behavior of elderly in daily life via their smart appliances at home. To the best of our knowledge, it is the first time to investigate the capacity of ML to analyze the human behavior (specifically for the group of elderly) by using the data recorded via smart appliances (television and refrigerator). The main contributions of this work can be summarized as: First, we present a novel ML framework for analyzing the daily behavior of elderly based on the data recorded in smart appliances, which can be easily implemented at home. Second, we systematically investigate and compare the ML paradigms, i.e., the classic ML models (with human handcrafted features and statistical models) and the state-of-the-art deep learning

(DL) models (with limited human involvement). Third, we aim to reveal the underlying mechanism of our proposed models and point out future directions of this domain, which we believe to benefit future studies.

The remainder of this article is organized as follows. First, we introduce the related work and background in Section II. Second, we describe the database and the methods used in Section III. Then, the experimental results and discussion are presented in Sections IV and V, respectively. Finally, we give the conclusion in Section VI.

II. Related Work and Background

As indicated by Sezer *et al.*, understanding the "context," i.e., making sense of the environment, situation, or status from sensor data, and then acting in an autonomous way, is critical for intelligent IoT [19]. In particular, understanding, learning, and reasoning from big data is paramount for the future success of IoT [19]. It has been shown that, the advancement of IoT technologies can benefit smart health monitoring in many aspects [20]. Among the previous studies, computational human behavior analysis (CHBA) plays an important role in developing the AIoT healthcare applications for its context-aware capacity to reflect the human's status in daily life. The relevant studies can be referred to as ambient assisted living (AAL) technologies and/or context-aware applications, which are well documented in [21] and [22]. Several review articles systematically and comprehensively summarized the main research topics and directions involved for AIoT-based elderly applications. Debes *et al.* [21] investigated the state-of-the-art sensor technologies and computational intelligence approaches for monitoring the field of activities of daily living (ADLs), which is a crucial part of AAL. They indicated that, the hybrid generative/discriminative methods, e.g., Fisher kernel learning (FKL), relying on kernel metric distances, are superior over traditional generative methods, e.g., hidden Markov models (HMMs). Moreover, they pointed out that the main challenge from a signal processing (SP) and ML side remains the generalizability over households. Mshali *et al.* [22] provided an overview of the most important functions and services offered by health monitoring systems (HMSs) for monitoring and detecting of human behavior. The authors painted a consolidated picture of not only the hardware architectures but also the computational algorithms and approaches in terms of context-aware applications in healthcare. As stated in their article, *context information*, refers to extracting high-level information, such as behavior patterns, or a subject's activity from the raw data collected from the sensors. Nathan *et al.* [23] further illustrated the data modalities and system paradigms for smart home applications for the ageing population. More recently, Deep *et al.* [24] introduced the concept of a dense sensing network (DSN), and analyzed its pros and cons for elderly anomaly behavior detection. They suggest that, when developing a reliable and robust elderly care system, three factors should be taken into account.

1) The system should be robust to environment changes.
2) The system should maintain the user's privacy.
3) The system should be convenient to use.

Additionally, the authors claim that a multisensor approach can achieve an impressive result by describing an activity in a comprehensive way from the sensors deployed in an indoor environment.

Ravì *et al.* proposed a method combining the features learnt from deep neural networks and human handcrafted features to analyze the human activities recorded via inertial sensors (e.g., accelerometers and gyroscopes) in [25]. They found that, a combination of deep learnt features and human handcrafted features can be better than the performance achieved by only using one of the aforementioned two. Venkatesh *et al.* [26] introduced a context-aware IoT system in smart city environment, which focused on analyzing the people-centric context in terms of user presence, user activity, air quality, and location data provided from sensors. Bianchi *et al.* [27] made a comparative study on different DL architectures for the human activity recognition task. They found that, a convolutional neural network (CNN) with four convolutional layers can reach the highest performance in classifying the human activities from WiFi wearable sensor data (3-D output from accelerometer, gyroscope, and magnetometer) [27]. Additionally, more kinds of smart wearable devices, like multichannel surface electromyography (sEMG) signals [28], skeleton joint sensors [29], or electroencephalography (EEG) [30], were demonstrated to be efficient to contribute to the AAL applications by leveraging the power of IoT and ML. More specifically, some AIoT technologies have been successfully applied for the ageing society.

In general, these methods can be divided into two categories: 1) sensor-based or 2) image-based models. For sensor-based models, we can see wearable sensors or nonwearable sensors applied to this field. When looking at the wearable sensor applications, the characteristics of temporal time series (TTSs) data have been considered in [31], where a hybrid architecture of a recurrent neural network (RNN) and a DNN was used for classifying the inputs, which have been reduced in dimension from the CNN. Moreover, Xu *et al.* [32] found that by introducing the activity similarity, the model's performance can be improved in elderly home activity recognition. Furthermore, Awais *et al.* [33] made a comprehensive study comparing the sensor locations of the human body for elderly physical activity classification (PAC). They find that a two-sensor solution (lower back plus thigh) can achieve the best classification performance. By surveying the nonwearable sensor studies, we can see that, Gochoo *et al.* [34] proposed a deep CNN (DCNN) for recognizing elderly individuals' travel patterns (direct, pacing, lapping, or random) from device-free nonprivacy invasive binary (passive infrared) sensor data. Furthermore, their DCNN model was demonstrated to be efficient in recognizing the activity images generated from the binary sensory data collected from the elderly subjects' home [35]. A combination of wearable and nonwearable sensors was also studied. Tsirmpas *et al.* [36] proposed a method for profile generation in an IoT environment, which can benefit AAL (e.g., for handicapped or elderly individuals living longer in their preferred environment). In their method, a self-organizing map (SOM) [37] and the fuzzy C-means (FCMs) [38] algorithm were used to model the users'

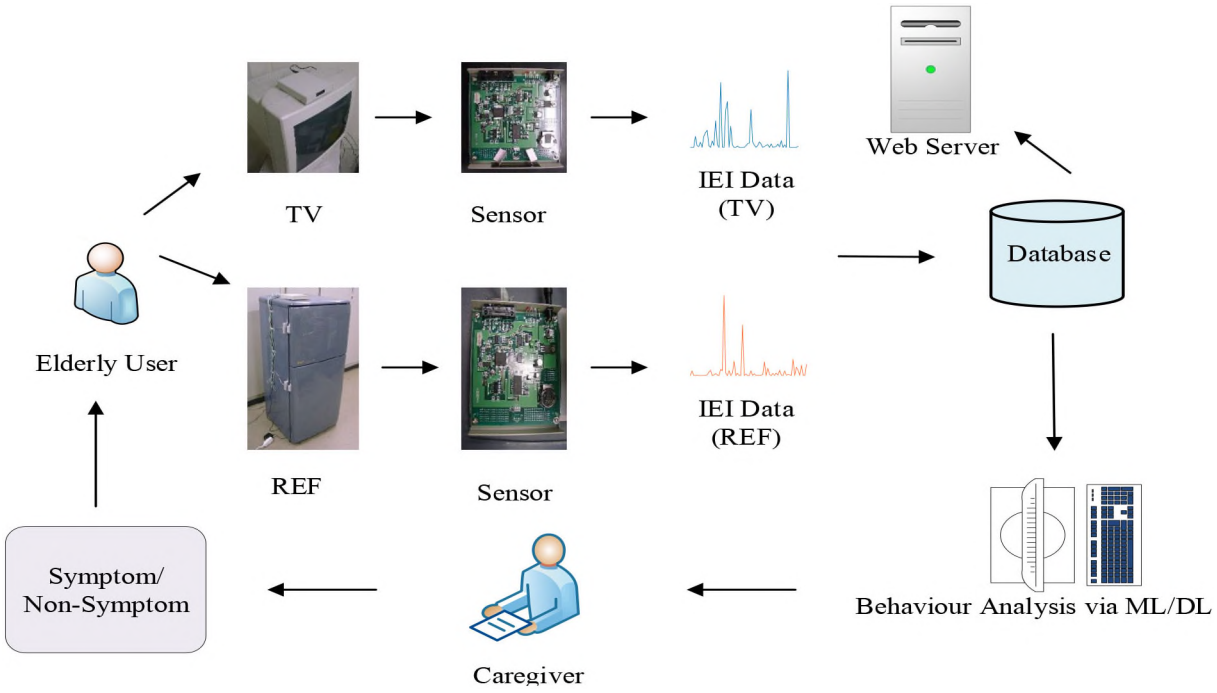


Fig. 1. Overview scheme of the proposed system. The IEI data inherited with the elderly spontaneous activity regulations are recorded via the home appliances, e.g., TV or refrigerator. By leveraging the power of ML or DL models, the caregiver can understand and make predictions on symptom or nonsymptom behaviors of the elderly individuals. Then, the elderly user can be supplied with healthcare services.

activities and their correlation with the available sensor. Their results were encouraging: high-level activities containing contextual information can be assembled by low-level activities recorded by raw sensor data (e.g., accelerometer), whereas, the underlying mechanism between the models and the human behavior was not fully discovered [36].

For image-based models, Hbali *et al.* [39] proposed a 3-D depth sensor (e.g., the Microsoft Kinect)-based low-cost system for elderly activity recognition. In their study, an extremely randomized trees (ERT) algorithm [40] was used for classifying the 3-D skeleton-based features. Furthermore, Chen *et al.* [41] introduced an activity encoding method that can convert skeleton sequence data to spatial-temporal images, which can be used for CNN-based feature extraction for elderly activity recognition. They successfully demonstrated their method in noisy data environment. Daubechies claim that, the Hilbert–Huang transformation (HHT) [42] can be superior to wavelet transformation (WT) [43] in providing sharp frequency resolution from Kinect sensor data [44].

Even though these pioneering work contributions showed promising results and potential, there are still two main limitations: First, these previous studies were not specifically designed for the elderly. In real practice, the inconvenience of elderly to make regular sports or activities as younger people should be taken into account. Therefore, some kinds of passive sensors (e.g., smart appliances) could be an alternative in this case. Second, more attention had been given to monitor the physical status of the subjects rather than the mental or psychological condition of the individual, which is also an important issue among elderly that cannot be ignored. A recent report indicated that, unstable marital status, unemployment, depressive symptoms, and mental disorder are independent risk factors for suicide in rural elderly [45]. To this end, we aim to propose a ML-based framework to detect the symptomatic days (associated with both their physical and psychological status) via their usage frequency recorded in smart appliances (i.e., refrigerator and television). A previous study of our group showed the statistical association between appliance usage frequency data and subjective symptoms of the elderly in [46]. We make a further investigation of that study by introducing the state-of-the-art techniques in SP and ML to the field of elderly in-home monitoring applications.

III. Database and Methods

In this section, we will first introduce the database for this study. Then, we describe the definition of the task and provide information of the data partitioning. Afterward, we give details about the proposed method. Fig. 1 describes the diagram of the proposed system in this study.

A. Interevent Interval Database

In this study, we apply the interevent interval (IEI) database recording the usage of two home appliances, i.e., television (TV) and refrigerator (REF) by elderly users—all of them are living alone in this study—in their daily life [46]. The environment for data collection is the home where the subjects stay. For the TV, the data acquisition device was equipped on the TV that can receive an infrared ray from a remote-control device, which recorded the time (in seconds) when the subject pushed any button. For the REF, a magnetic switch recorded the time when the door was opened or closed by the subject.

Therefore, these IEIs can describe the behavior patterns of the usage of the appliances by the elderly in their daily life. For details of the data collection process, it can be referred to [46]. In the following parts of this article, we name the IEI data for the TV and the REF as IEI-TV and IEI-REF, respectively.

1) Data Preprocessing: The original IEI data were recorded over 31 consecutive days per participating subject [46]. Due to the setup and removing of the devices, the IEI data recorded in the first and the last day were excluded in our study as the same in [46]. Furthermore, we excluded the IEI data which has a limited number (less than 32, see [46]) of the IEIs to guarantee the ML models can capture the changes over a sufficient number of IEIs. Additionally, to eliminate the effects by night sleep, the first and the last of the per-day IEI were excluded. Finally, considering usage in future studies and an in-depth analysis of the subject behavior, we selected those subjects who have both qualified IEI-TV and IEI-REF data, which results in a number of 76 subjects (female: 52, male: 12, unknown: 12) from the original 100 subjects. The age range of these selected 76 subjects is from 70 to 98 within a mean value of 75.4 ($\pm$4.9).[1]

2) Ground Truth Annotation: All the subjects had been asked to provide a self report, including a series of questions about their mood, appetite, and sleep quality when getting up in the morning [46]. In this study, based on initial experiments, we selected three subjective feelings (*appetite*, *pain*, and *sleep quality*) as basic measure to match the ground truth of the IEI data. We can see that, the status we take into account, contains not only the physical health (e.g., appetite) but also the mental health (e.g., sleep quality). In this study, our task is to match the behavior of elderly in daily life to their *Symptom* or *NonSymptom* date during a long-term monitoring. Therefore, we combine the self report and their answers to annotate the IEI data, which can be found in Table I. Our target is to detect out of the *Symptom* instances as many as possible; we define this kind of instances as: "Bad" for appetite, or "Yes" for pain, or Bad for sleep quality. Considering future applications in the real-world, we hope that such an elderly care system can be sensitive enough for emergency events. Note that the definitions of the *Nonsymptom* instances are very rigid and coarse: Only "Good" or "Normal" for appetite, and "No" for pain, and Good or Normal for sleep quality. Fig. 2 shows the examples of the IEI data annotated as "Symptom" or "Nonsymptom."

3) Data Partition: In order to avoid over-optimistic results, we take *subject-independency* into account, i.e., the IEI data for training and testing the ML models are rigidly stemming from different subjects. All the 76 subjects' IEI data are split into three sets, i.e., *train* (46 subjects, 60 % of overall), *development* (15 subjects, 20 % of overall), and *test* (15, 20 % of overall). The hyperparameters of the ML models will be tuned and optimized on the train and development sets. Then, the test set (unseen) will be used to validate the final performance by the ML models trained on the train plus development sets with the optimized hyperparameters. Table II shows the instance numbers of each data set in our study.

[1]Among these statistics, 12 subjects who did not share all information were not included.

TABLE I
QUESTIONNAIRE-BASED SELF REPORT AND THE GROUND-TRUTH ANNOTATION. (a) QUESTIONNAIRE. (b) ANNOTATION

(a)

Status	Answers
Appetite (**A**)	1: Good; 2: Normal; 3: Bad.
Pain (**P**)	1: Yes; 2: No.
Sleep Quality (**SQ**)	1: Good; 2: Normal; 3: Bad.

(b)

Annotation	Condition
Symptom	**A** (3) $\|\|$ **P** (1) $\|\|$ **SQ** (3)
Non-Symptom	**A** (1 $\|\|$ 2) & **P** (2) & **SQ** (1 $\|\|$ 2)

B. Machine Learning Paradigms

In this study, we mainly focus on investigating and comparing two ML paradigms, i.e., training an ML model with human handcrafted features, and training a DL model directly using the raw sensor data without any human expert domain knowledge. For the former paradigm, we use two methods, statistical functionals (func.) and bag-of-behavioral-words (BoBW). For the latter one, we use an end-to-end (e2e) deep architecture.

1) Statistical Functionals: The analysis of the change of the IEI data over a given time period is essential for further ML model building. In addition, for the static classifiers, e.g., a support vector machine (SVM) [47], we should extract superasegmental features [48] that are independent of the length of the analysed data. Motivated by our previous work [49], we applied nine statistical functionals successfully used in human behavior analysis. These functionals (see Fig. 3) include *maximum*, *minimum*, *mean*, *range*, *standard deviation*, *skewness*, *kurtosis*, *slope*, and *bias* of linear regression approximation, which are extracted from the IEI data recorded per day. We found the selected functionals can be efficient and robust for describing the fluctuations of the time-series signals over a given time period. For instance, the extreme values (e.g., maximum, minimum) can represent the elderly behavior changes in a time duration. The skewness and kurtosis can be good indicators to reflect the signal distribution, which may carry important information about the behavior regulation. The linear regression estimators (the slope and bias) can describe the trend of the time-series signals, which may reflect the behavior changes in a specific duration.

Let $y = x(i),\ i = 1, 2, \ldots, N$ denote a sequence of the IEI data in a given time period. Then, the maximum ($x_{\max}$), minimum ($x_{\min}$), mean (μ_x), range (λ_x), standard deviation (σ_x), skewness ($\hat{s}$), and kurtosis ($\hat{k}$) values are defined as

$$x_{\max} = \max\{x(1), x(2), \ldots, x(N)\} \tag{1a}$$

$$x_{\min} = \min\{x(1), x(2), \ldots, x(N)\} \tag{1b}$$

$$\lambda_x = x_{\max} - x_{\min} \tag{1c}$$

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x(i) \tag{1d}$$

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} |x(i) - \mu_x|^2} \tag{1e}$$

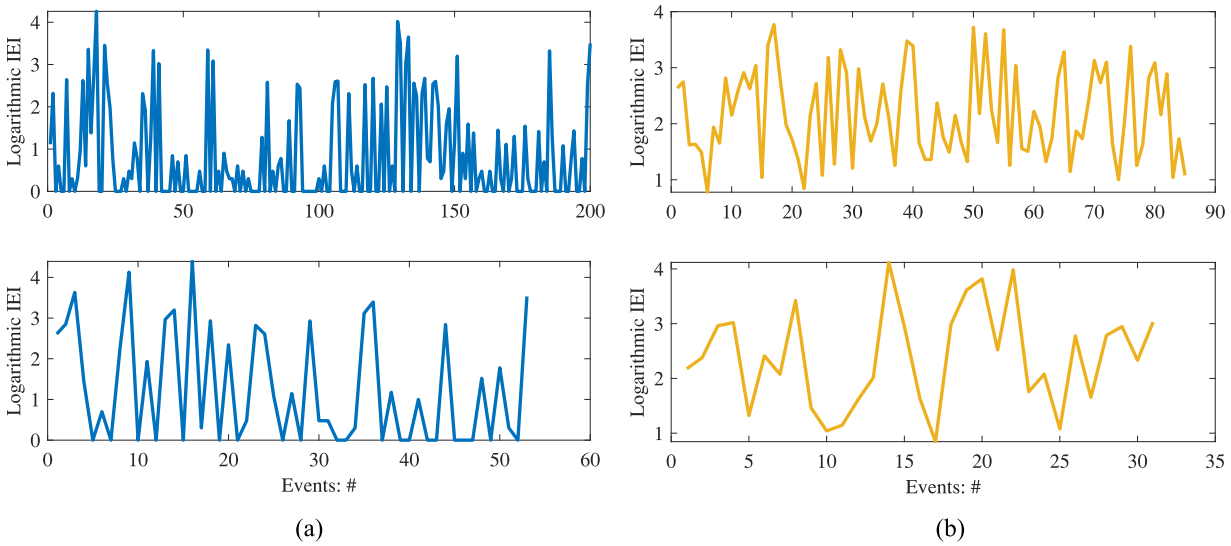


Fig. 2. Examples of IEI data samples (IEI data unit: second). The IEI samples annotated as "Symptom" show more frequent changes in one day than the "Nonsymptom" samples. (a) IEI-TV Examples (Top: Symptom; Bottom: Nonsymptom). (b) IEI-REF Examples (Top: Symptom; Bottom: Nonsymptom).

TABLE II
NUMBER OF INSTANCES IN EACH DATA SET. THE DATA PARTITION OF SUBJECTS ARE THE SAME FOR BOTH THE IEI-TV AND THE IEI-REF DATA WHILE THEY HAVE DIFFERENT INSTANCE NUMBERS. (a) IEI-TV DATA. (b) IEI-REF DATA

(a)

	Train	Dev	Test	Σ
Symptom	223	110	100	433
Non-Symptom	493	136	161	790
Total	716	246	261	1 223

(b)

	Train	Dev	Test	Σ
Symptom	217	99	93	409
Non-Symptom	438	101	110	649
Total	655	200	203	1 058

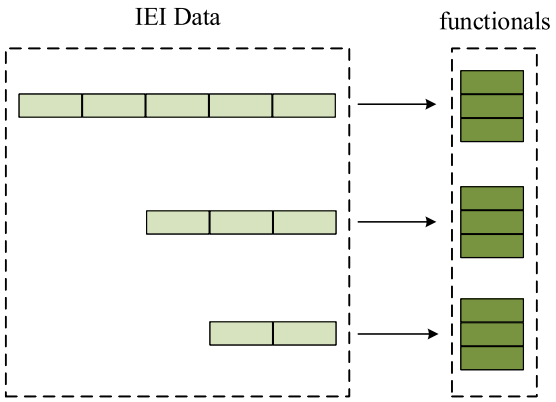


Fig. 3. Diagram of the functionals approach. Independent of the length of the instances, *max*, *min*, *mean*, etc., are extracted from the IEI data.

$$\hat{s} = \frac{E(x-\mu_x)^3}{\sigma_x^3} \tag{1f}$$

$$\hat{k} = \frac{E(x-\mu_x)^4}{\sigma_x^4} \tag{1g}$$

where $E(\cdot)$ denotes the expected value. For the linear regression, the target is to approximate a line ($\tilde{y} = \alpha i + \beta$) that has the minimized *quadratic error* ($\hat{e}^2$) between the approximated line ($\tilde{y}$) and the actual value series (y). $\hat{e}^2$ can be written as

$$\begin{aligned}\hat{e}^2 &= \sum_{i=1}^{N}(y-\tilde{y})^2 = \sum_{i=1}^{N}(x(i)-\alpha i-\beta)^2 \\ &= \sum_{i=1}^{N}\Big(x(i)^2 - 2\alpha i x(i) - 2\beta x(i) + 2\alpha\beta i + \alpha^2 i^2 + \beta^2\Big)\end{aligned} \tag{2}$$

where α and β represent the slope and the bias, respectively. To minimize $\hat{e}^2$, the following differential equations should be applied (see [48]):

$$\frac{\partial \hat{e}^2}{\partial \alpha} = \sum_{i=1}^{N}\Big(-2ix(i) + 2\beta i + 2\alpha i^2\Big) = 0 \tag{3a}$$

$$\frac{\partial \hat{e}^2}{\partial \beta} = \sum_{i=1}^{N}(-2x(i) + 2\alpha i + 2\beta) = 0 \tag{3b}$$

which can be rewritten as

$$-\sum_{i=1}^{N} ix(i) + \beta\sum_{i=1}^{N} i + \alpha\sum_{i=1}^{N} i^2 = 0 \tag{4a}$$

$$-\sum_{i=1}^{N} x(i) + \alpha\sum_{i=1}^{N} i + N\beta = 0. \tag{4b}$$

Then, the solutions for α and β can be yielded as

$$\alpha = \frac{N\sum_{i=1}^{N} ix(i) - \sum_{i=1}^{N} i\sum_{i=1}^{N} x(i)}{N\sum_{i=1}^{N} i^2 - \Big(\sum_{i=1}^{N} i\Big)^2} \tag{5a}$$

$$\beta = \frac{\sum_{i=1}^{N} x(i)\sum_{i=1}^{N} i^2 - \sum_{i=1}^{N} i\sum_{i=1}^{N} ix(i)}{N\sum_{i=1}^{N} i^2 - \Big(\sum_{i=1}^{N} i\Big)^2}. \tag{5b}$$

Fig. 4 shows the 3-D scatter plots of the selected three functionals via ReliefF algorithm [50] for IEI-TV and IEI-REF

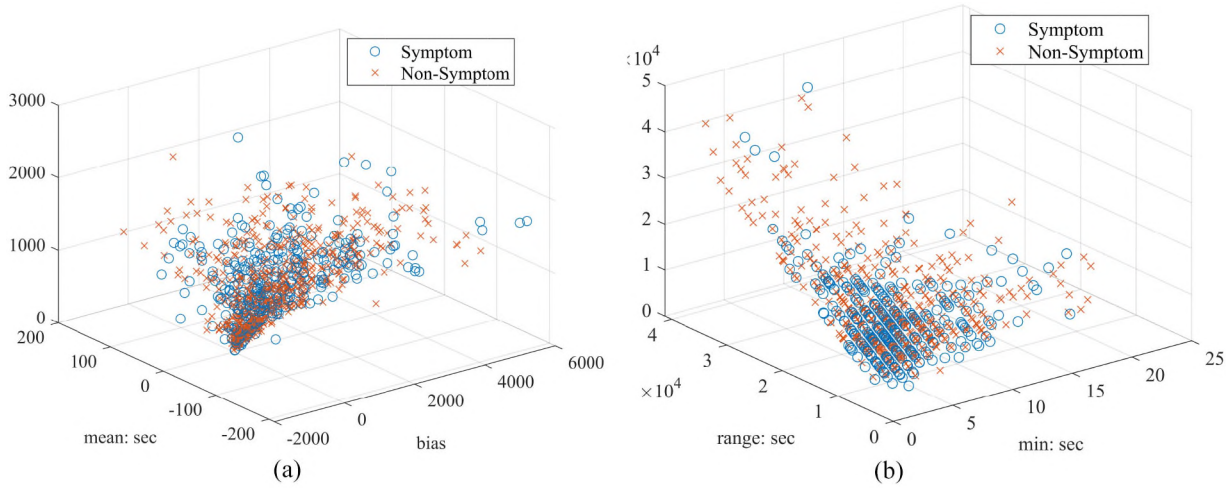


Fig. 4. Visualization of the 3-D scatter of selected IEI functionals in the train plus dev sets. The functionals are top three features (ranked by their contributions to the binary classification task) selected via the ReliefF algorithm [50]. The top three functionals for IEI-TV are bias, slope, and mean values. The top three functionals for IEI-REF are min, max, and mean values. (a) 3-D Scatter of selected IEI-TV functionals. (b) 3-D scatter of selected IEI-REF functionals.

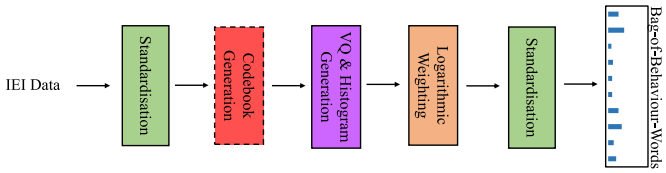


Fig. 5. Processing chain of the bag-of-behavior-words (BoBW) approach.

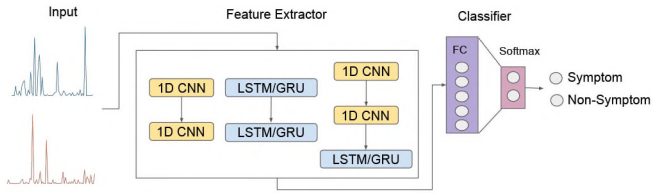


Fig. 6. Diagram of the e2e method. In this paradigm, higher representations can be learnt automatically from a combination of CNNs and/or RNNs (with LSTM or GRU cells).

Raw wave 150 frames
4@4, *BN*
4@8, *BN*
FC 2048, ReLU
FC 2, Softmax

(a)

Raw wave 10 frames × 29 sequences
LSTM 50 hidden nodes
LSTM 50 hidden nodes
FC 2048, ReLU
FC 2, Softmax

(b)

Raw wave 150 frames
4@3, *BN*
4@6, *BN*
LSTM 50 hidden nodes
LSTM 50 hidden nodes
FC 2048, ReLU
FC 2, Softmax

(c)

Fig. 7. Architectures of the (a) CNN, (b) RNN, and (c) CNN–RNN for our proposed e2e method. BN: Batch Normalization; FC: Fully Connected; ReLU: Rectified Linear Unit. The LSTM unit can be replaced by a GRU one determined by the optimization process on the dev set.

data. We can see that, the 3-D feature map cannot sufficiently distinguish the two groups of IEI for both the TV and the REF IEI data. More features are needed to train an efficient classifier. In this study, we investigate and compare eight ML models using functionals, which include naïve Bayes (NB) [51], linear discriminant analysis (LDA) [52], k-nearest neighbor (*k*-NN) [53], random forest (RF) [54], support vector machine (SVM) [47], extreme learning machine (ELM) [55], kernel-based extreme learning machine (KELM) [56], and a deep neural network (DNN) [57].

2) Bag-of-Behavior-Words: The BoBW approach originated from the bag-of-words (BoW) principle, which appeared early in a description written in [58]. In the past decade, the BoW approach has been successfully applied to the field of *computer vision* [59], *natural language processing* [60], *acoustic event classification* [61], *speech emotion recognition* [62], and *health care* [63], among others. Most recently, we introduced the BoW approach into the analysis of human behavior, and named it as BoBW [64].

In this study, the IEI data (frames with a window length of 10 points and an overlap of 5 points) are the inputs of the BoBW paradigm (see Fig. 5). At first, the IEIs will be passed through a process called vector quantization (VQ). This process is completed by using a *codebook* containing template IEIs which is previously learnt from a certain amount of the training data.

For generating the codebook, we use a *random sampling* strategy [61] of the IEIs that following the initialization step of *K-means++ clustering* [65] instead of the classic *K-means clustering* [66], [67]. Compared to the classic *K-means clustering*, *K-means++ clustering* allows for a better initialization step, which is independent of the initial centroid. Then, the N_a words (the IEI templates) with the lowest *Euclidean* distance will be considered when assigning each IEI to the generated codebook. Finally, the logarithm (with a bias of one) is taken from the word frequencies to compress the range of values.

3) End-to-End Learning: The e2e models utilize the capacity of DL to learn higher representations from the raw data via a series of nonlinear transformations. In this paradigm (see Fig. 6), human handcrafted features (e.g., functionals) are not needed. Previous studies have demonstrated the efficiency of e2e models among others in music analysis [68], speech emotion recognition [69], and healthcare [70]. In this study,

we investigate and compare several topologies of e2e models, which include not only using CNN [71] or RNN [72] but also a combination of the two aforementioned typical DL models. For RNN models, we use both long short-term memory (LSTM) [73] and gated recurrent unit (GRU) [74] cells to overcome the *vanishing gradient* problem in RNN training [75]. In particular, we also investigate the bidirectional RNNs [76] for achieving more contextual information (both the previous and future contextual information) of the sequence of IEI rather than unidirectional RNNs. The CNN was demonstrated to extract spatial and temporal features [35] while the RNN was able to capture the time-series characteristics [31]. A hybrid architecture composed of CNN and RNN has been demonstrated to excel in our previous e2e learning work [69]. We designed the architectures of the CNN and RNN models based on empirical settings in our initial experiments. Fig. 7 shows the candidate architectures of our e2e models. In this study, the *binary cross entropy* is used as the loss function.

C. Evaluation Metrics

To evaluate the performance of the proposed methods in this study, we use unweighted average recall (UAR) as the main metric together with some other frequently used metrics found in binary classification tasks.

1) Unweighted Average Recall: Considering the imbalanced data distribution given in the data, we choose UAR as the main metric in this study. Compared to weighted average recall (WAR, i.e., accuracy), UAR is the averaged recall achieved for each class [77], [78], which can avoid over-optimistic results caused by using WAR due to over-training to one class of the data which has a larger proportion than others of the overall data set.

2) Complementary Metrics: Additionally, we use *accuracy* (Acc.), *sensitivity* (Sens.), *specificity* (Spec.), *precision* (Prec.), *F1-score*, and *geometric mean* (G-mean), as the complementary metrics.

IV. Experimental Results

We will present the experimental results in this section. The detailed experimental setup will be given followed by the results achieved by the methods proposed previously.

A. Setup

All the proposed ML/DL models are optimized (with hyper-parameters) by a grid-search strategy (see Table III) on the development set. Then, the optimized models are validated by applying the optimized hyperparameters on the test set. In the following experiments, results are shown on the best results achieved on the development set, and the ones reached by the optimized models on the test set. The models using functionals are implemented by MATLAB R2019a (by MathWorks) except for SVM which are given by the LIBSVM [79] toolkit. The BoBW approach is provided by the openXBOW [80] toolkit, while the e2e models are developed on Python 3.7.5 and PyTorch 1.3.0. To eliminate the effects by outliers, all the data are standardized before being fed into the models.

TABLE III
Grid-Search Process for Optimizing the Hyperparameters of the Models. These Hyperparameters Are Optimized on the Development Set, and Applied to the Test Set. RBF: Radial Basis Function

Model	Main Hyper-parameters
NB	*kernel smoothing density estimate*: 'normal', 'box', 'epanechnikov', 'triangle'.
LDA	*discriminant type*: 'linear', 'diaglinear', 'pseudolinear'; γ *values*: 0.1, 0.2, 0.3, $\cdots$, 1.0
k-NN	k *values*: 1, 10, 20, $\cdots$, 90, 100; *distance metrics*: 'cityblock', 'chebychev', 'correlation', 'cosine', 'euclidean', 'hamming', 'jaccard', 'minkowski', 'seuclidean', 'spearman'.
RF	*number of trees*: 10, 50, 100, 200, 400, 600, 800, 1 000, 2 000, 3 000; *fraction for the treebagger*: 0.1, 0.2, 0.3, $\cdots$, 1.0.
SVM	*kernels*: 'linear', 'polynomial', 'RBF', 'sigmoid'; C *values*: 10^{-5}, 10^{-4}, $\cdots$, 10^{4}, 10^{5}.
ELM	*activation functions*: 'sigmoid', 'sine', 'hardlim' 'tribas', 'radbas'; *number of hidden units*: 10, 50 100, 200, 500, 1 000, 3 000, 5 000, 8 000, 10 000 C *values*: 10^{-5}, 10^{-4}, $\cdots$, 10^{4}, 10^{5}.
KELM	*kernels*: 'linear', 'polynomial', 'RBF', 'wavelet'; C *values*: 10^{-5}, 10^{-4}, $\cdots$, 10^{4}, 10^{5}.
DNN	*four* hidden layers; *optimiser*: 'trainscg'; *number of hidden units*: 32, 64, 128, 256, 512, 1 024
CNN	*learning rate*: 10^{-3}; *channels*: 3-6, 4-8; *kernel size*: 8-4, 4-4; *stride size*: 8-4, 4-4;
LSTM	*learning rate*: 10^{-3}; *direction*: 'uni-direction', 'bi-direction'; *hidden layers*: 10, 50; *hidden nodes*: 1, 2;
GRU	*learning rate*: 10^{-3}; *direction*: 'uni-direction', 'bi-direction'; *hidden layers*: 1, 2; *hidden nodes*: 10, 50;

TABLE IV
UARs (in [%]) Achieved by ML Models Using Functionals. The Results on the Dev Set Are Achieved by the Optimal Hyperparameters. The Best Results on the Test Set Are Highlighted in Bold Face. Chance Level: 50.0 %. (a) IEI-TV. (b) IEI-REF

(a)

	NB	LDA	k-NN	RF	SVM	ELM	KELM	DNN
Dev	51.6	50.5	56.4	54.9	55.9	57.3	59.7	50.9
Test	52.7	**52.8**	49.7	50.2	52.0	52.1	50.5	51.9

(b)

	NB	LDA	k-NN	RF	SVM	ELM	KELM	DNN
Dev	52.4	55.9	57.9	57.3	57.4	58.8	56.6	54.7
Test	49.6	50.0	48.6	51.4	52.5	46.2	46.8	**54.4**

B. Results

The results (UARs in [%]) of the proposed models are illustrated in Tables IV–VI, respectively.

For the models trained by functionals, the best performances on the test set by TV and REF data are reached by LDA (a UAR of 52.8 %) and the DNN (a UAR of 54.4 %), respectively. For the models built on BoAW representations and SVM, the best performances on the test set by the TV and the REF data obtain a UAR of 54.3 % and 58.7 %, respectively. For the e2e models, best performances on the test sets of TV and REF

TABLE V
UARs (IN [%]) ACHIEVED BY A SVM MODEL USING BoBW. THE RESULTS ON THE DEV SET ARE ACHIEVED BY THE OPTIMAL HYPERPARAMETERS. THE BEST RESULTS ON THE TEST SET ARE HIGHLIGHTED IN BOLD FACE. C_s: CODEBOOK SIZE; N_a: ASSIGNMENT NUMBER. CHANCE LEVEL: 50.0 %. (a) IEI-TV. (b) IEI-REF

(a)

	C_s =	10	50	100	500	1 000
N_a = 1	Dev	50.9	53.1	56.5	52.2	56.1
	Test	49.9	44.6	45.3	**54.3**	51.2
N_a = 5	Dev	51.5	54.1	52.8	54.2	55.4
	Test	49.6	47.3	49.1	47.0	52.7
N_a = 10	Dev	50.0	56.9	55.9	52.4	51.8
	Test	50.0	50.2	51.8	46.6	48.5

(b)

	C_s =	10	50	100	500	1 000
N_a = 1	Dev	52.3	52.4	58.3	60.4	51.5
	Test	51.0	51.8	53.2	56.8	**58.7**
N_a = 5	Dev	56.0	55.8	51.3	59.9	54.3
	Test	51.0	52.3	47.7	47.4	47.0
N_a = 10	Dev	50.4	57.8	57.3	55.8	54.7
	Test	45.0	52.2	50.4	49.1	51.7

TABLE VI
UARs (IN [%]) ACHIEVED BY E2E MODELS. THE RESULTS ON THE DEV SET ARE ACHIEVED BY THE OPTIMAL HYPERPARAMETERS. THE BEST RESULTS ON THE TEST SET ARE HIGHLIGHTED IN BOLD FACE. CHANCE LEVEL: 50.0 %. (a) IEI-TV. (b) IEI-REF

(a)

	CNN	LSTM	GRU	CNN+LSTM	CNN+GRU
Dev	52.4	50.6	52.3	50.5	50.5
Test	49.7	46.9	**53.3**	50.0	50.0

(b)

	CNN	LSTM	GRU	CNN+LSTM	CNN+GRU
Dev	52.1	54.6	55.1	53.1	53.6
Test	51.2	**53.4**	49.2	51.6	52.5

data are 53.3 % and 53.4 % UAR, respectively. Fig. 8 shows the training and dev losses of the best e2e models.

Fig. 9 shows the complementary metrics (in [%]) as reached by the best models on the test set. We can see that, both the functional and BoBW models master a high specificity (more than 90.0 %) based on the TV data while their sensitivities are extremely low (lower than 20.0 %). The exception if the e2e model, which reaches a sensitivity of 79.0 % with a low specificity of 33.5 %. In contrast, when using REF data, the sensitivities can be considerably improved whereas a sacrifice in specificities occurs, and *vice versa*.

Additionally, when comparing the accuracies and precisions, the functional and BoBW based models using the TV data are found to be superior to their counterparts using the REF data, whereas the contrary phenomenon can be found when evaluating their F1-scores and G-means. For the e2e method, the TV data based model shows a higher precision, F1-score, and G-mean, but yields to the REF data based model in accuracy.

Table VII presents the confusion matrices of the best models on the test set. It can be seen that, most of the best models have

TABLE VII
CONFUSION MATRICES (NORMALIZED: IN [%]) BY THE BEST MODELS ON THE TEST SET. **S**: SYMPTOM; **N**: NONSYMPTOM. THE HYPERPARAMETERS OF THE MODELS ARE: 1) FUNC. (TV: LDA–*type*: "DIAGLINEAR," *g*-VALUE: 0; REF: DNN–[256-256-256-256].); 2) BoBW: (TV: $C_s = 500$, $N_a = 1$, SVM–*kernel*: "POLYNOMIAL," *C*-VALUE: 100; REF: $C_s = 1\,000$, $N_a = 1$, SVM–*kernel*: "SIGMOID," *C*-VALUE: 100.); 3) E2E: (TV: *e2e model*: "GRU," *hidden layers*: 2, *hidden nodes*: 50; REF: *e2e model*: "LSTM," *hidden layers*: 1, *hidden nodes*: 10.). (a) FUNC.: TV. (b) BoBW: TV. (c) E2E: TV. (d) FUNC.: REF. (e) BoBW.: REF. (f) E2E: REF

(a)

Pred ->	S	N
S	8.0	92.0
N	2.5	97.5

(b)

Pred ->	S	N
S	16.0	84.0
N	7.5	92.5

(c)

Pred ->	S	N
S	79.0	21.0
N	66.5	33.5

(d)

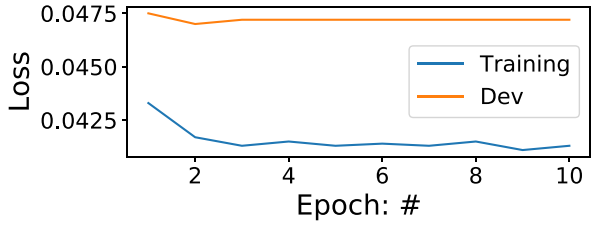
Pred ->	S	N
S	28.0	72.0
N	19.1	80.9

(e)

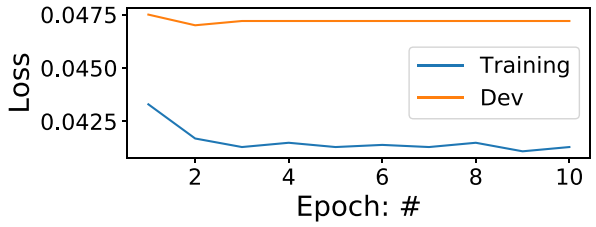
Pred ->	S	N
S	53.8	46.2
N	36.4	63.6

(f)

Pred ->	S	N
S	25.8	74.2
N	19.1	80.9



(a)



(b)

Fig. 8. Visualization of losses of the best two e2e models in the training and dev sets. The training losses are converged in 10 epochs. Then, the learnt weights of the epoch by which the dev loss is the minimized, is applied to the test set. (a) GRU–RNN for IEI-TV. (b) LSTM–RNN for IEI-REF.

a low recall on detecting the *Symptom* instances whereas a high recall on finding *Nonsymptom* instances. The e2e model using the TV data can have a high detection rate of the *Symptom* instances while it also tends to a high false alarm rate, i.e., more than 50 % *Nonsymptom* instances are incorrectly recognized as *Symptom* instances. The best model, i.e., the BoBW method using the REF data has a recall of 53.8 % on *Symptom* instances and a recall of 63.6 % on *Nonsymptom* instances.

V. DISCUSSION

In this section, we analyse and discuss the results achieved in the current study. Furthermore, we summarize the limitations and provide our perspectives for future work.

A. Current Findings

Even though the current results are limited in their performance, most of the best models in this study can beat the chance level (50.0 % in UAR) already given the fairly

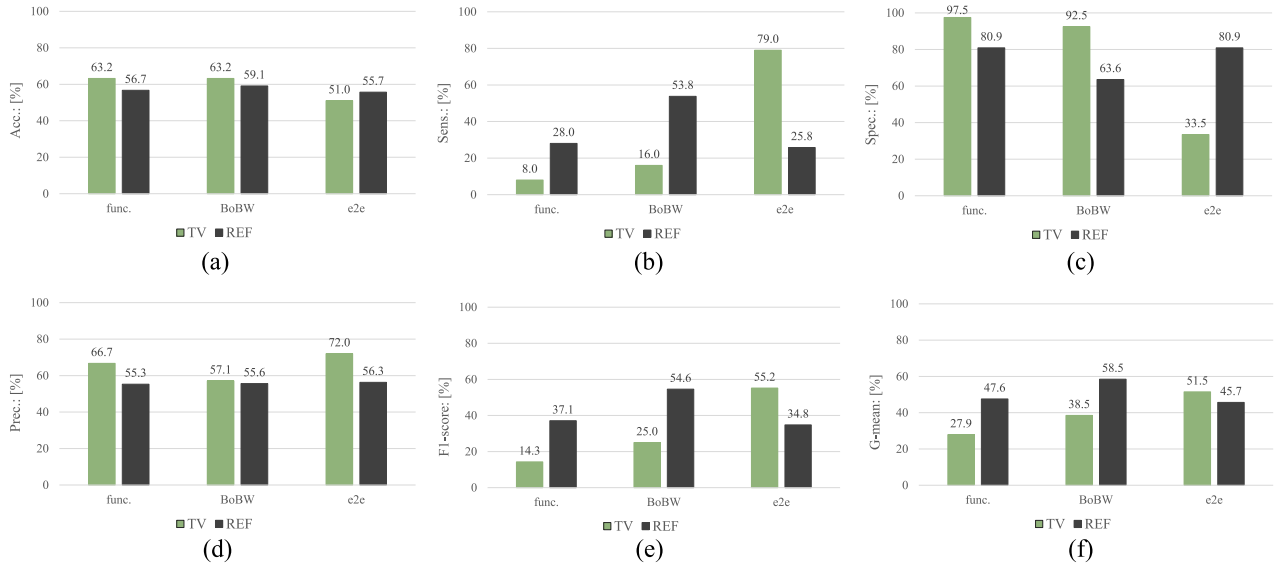


Fig. 9. Complementary evaluation metrics (in [%]) achieved by the best models on the test set. The hyperparameters of the models are: 1) func. (TV: LDA–*type*: "diaglinear," *g*-value: 0; REF: DNN–[256-256-256-256].); 2) BoBW: (TV: $C_s = 500$, $N_a = 1$, SVM–*kernel*: "polynomial," *C*-value: 100; REF: $C_s = 1\,000$, $N_a = 1$, SVM–*kernel*: "sigmoid," *C*-value: 100.); 3) e2e: (TV: *e2e model*: "GRU," *hidden layers*: 2, *hidden nodes*: 50; REF: *e2e model*: "LSTM," *hidden layers*: 1, *hidden nodes*: 10.). (a) Acc. (b) Sens. (c) Spec. (d) Prec. (e) F1-score. (f) G-mean.

TABLE VIII
COMPARISON BETWEEN STUDIES IN THE EXISTING LITERATURE ON AIOT FOR ELDERLY BEHAVIOR ANALYSIS. SOMS: SELF ORGANIZING MAPS; FCM: FUZZY C-MEANS (FCM); ERT: EXTREMELY RANDOMIZED TREES; DCNN: DEEP CONVOLUTIONAL NEURAL NETWORK; CNN: CONVOLUTIONAL NEURAL NETWORK; DNN: DEEP NEURAL NETWORK; LSTM–RNN: LONG SHORT-TERM MEMORY-BASED RECURRENT NEURAL NETWORK; IMA: INTERFRAME MATCHING ALGORITHM; RF: RANDOM FOREST; AS: ACTIVITY SIMILARITY; HHT: HILBERT–HUANG TRANSFORMATION; SVM: SUPPORT VECTOR MACHINE; CFS: CORRELATION-BASED FEATURE SELECTION; FCBF: FAST CORRELATION-BASED FILTER; PAC: PHYSICAL ACTIVITY CLASSIFICATION. † INDICATES THE WEARABLE SENSORS; ‡ INDICATES THE NONWEARABLE SENSORS

Ref.	Data	Methods	Pros	Cums
[31]	Sensors†	CNN, DNN, LSTM–RNN	efficient to detect Parkison's disease; low input dimensions	time-consuming when training; not validated in the wild
[32]	Sensors†	RF, AS	activity similarity can improve the model's performance	the type and scale of acquired sensor data was limited
[33]	Sensors†	CFS, FCBF, ReliefF, SVM	a detailed study on sensor locations, and feature selection had been done	a larger number of participants was lacking
[34]	Sensors‡	DCNN	early detection of dementia; non-privacy invasive	vulnerable for a few episodes of high number of movements; cannot be used when there is a visitor
[35]	Sensors‡	DCNN	can extract the intra-sensor patterns of the binary sensor data	temporal information of the data might be ignored
[36]	Sensors†,‡	SOMs, FCM	can handle inference for high-level and more complex activities	dependent on context and training data; estimation of the profiles has to be repeated in some cases
[39]	Images	Spatial and Temporal features, ERT	low-cost; real-time monitoring	the number of activities for the system to recognise is quite limited
[41]	Images	Activity Encoding IMA, CNN	robust to the environmental condition and the quality of data	dependent on the beginning of an activity; vulnerable for a target object disappearance by obstruction
[44]	Images	HHT, SVM	compared to wavelet analysis, HHT can give a sharper frequency resolution	features related to cognitive performance were not involved
Proposed	Sensors‡	Classic ML BoBW, e2e	novel modality of appliance usage data; validated by spontaneous real-world data	fundamental study of data characteristics is lacking; accurate objective groundtruth is lacking

basic approach considered by the technical home setup chosen. Among the proposed models, the BoBW approach outperforms the other two methods, i.e., functionals and e2e. We may think that, the BoBW approach can provide a global view of the whole training data when extracting statistical information, which makes it better suited than the functional approach (only focusing on one instance). This finding is consistent with our previous study on using BoBW to address the major depressive disorder detection task [64]. The e2e models are demonstrated to show their potential in describing elderly behaviors without any human expert domain knowledge. More specifically, e2e models can be comparable to functional (depending on human handcrafted features) models.

For the functional models, there appear no big gaps between each ML model (see Table IV). For most of the models, there is a decrease of performance on the test set comparing to the

development set. This could be due to the overfitting caused by the limited size of the database. The DNN model, as the best one using the REF data by functionals, shows the smallest gap between the development and the test set (54.7 % versus 54.4 %).

For the BoBW (plus SVM) models, we find that C_s and N_a are crucial hyperparameters for final performances. In this study, the best performances by the TV and the REF data are all using the smallest number of assignments, i.e., $N_a = 1$. However, larger number of codebook sizes (e.g., 500, 1000 for TV and REF, respectively) may help reach higher UARs by using this smallest assignment number ($N_a = 1$). We need to note that, compared to the functional method, the BoBW requires considerably less human knowledge. In this study, we use the raw IEI data as the inputs of the BoBW processing chain, which can learn higher representations (i.e., histograms) automatically in an unsupervised scenario.

For the e2e models, the best models based on the TV data and the REF data are all achieved by the RNN models, namely, the GRU–RNN and the LSTM–RNN models, respectively. There are no considerable improvements when combining the RNN and the CNN models. We think that, the time series of using the TV and the REF may be sufficient to capture the contextual information which can describe the state of an elderly home appliance user in their daily life. The capacity of CNNs to extract features from the raw data should be further investigated in future work. The LSTM–RNN model reaches the best performance (a UAR of 53.4 %) using the REF data. We think this may be benefited from a strong relation between the appetite of a user and the usage of the refrigerator (see Table I).

A comparison between the proposed method and the state-of-the-art previous works on AIoT for elderly behavior analysis is shown in Table VIII. Due to the factor that the data modality, methods, and evaluation metrics are varied among different studies, we cannot make a direct comparison of the results. Nevertheless, we may summarize as follows.

1) Most of the existing studies were based on the data collected from a lab environment, which means the subjects (elderly individuals) may not perform their activities in a spontaneous scenario. At this point, our proposed method is validated on the data collected from real-world spontaneous activities in the elderly participants' daily life.
2) Most of these previous studies need to reset the whole system (including both hardware and software) if the environment is changed. In contrast, our proposed method can be adapted easily and flexibly to a new environment. In fact, the TV and refrigerator are all prevalent appliances.
3) Our study takes subject independency into account, which is ignored by some of the previous studies. This supports the provision of a reasonable result rather than an over-optimistic number.

B. Limitations and Outlook

Data scarcity (annotation) is still a serious challenge in this or other relevant studies. On the one hand, we can easily collect IEI data with the prevalent smart appliances—not only the TV or REF types but also some others, such as based on the usage of lights, the microwave, or the air-conditioner. On the other hand, accurate annotations are still lacking, which extremely restrains the current performance of the models, specifically for the DL models. In addition, data imbalance characteristics cannot be ignored. In this study, the *Nonsymptom* instances have a larger proportion in the total database than the *Symptom* instances. In future work, we should consider *data augmentation* techniques to overcome the aforementioned issues. In this study, we tried simple methods like *data upsampling* and *time shifting*. But the improvement is not considerable or stable. We will investigate more advanced approaches like generative adversarial networks (GANs) [81].

Moreover, the current annotations are made by participants' self-reports, which maybe very subjective. In the future, we can consider using more reasonable and objective annotation methods as in our previous work on *spontaneous physical analysis* [49], [64], [82]. Accurately annotating the status of the elderly is a difficult, but essential work for ML-based methods.

In addition, we only studied the two data modalities (i.e., IEI-TV and IEI-REF) separately in this work. The reason is due to the limited condition that, we cannot find sufficiently overlapped instances (with accurate annotation in each corresponding date) of the two modalities. In future work, one should investigate a combination of the two modalities, which may improve the final prediction performance of the model. Also, studying the relationship between the two modalities may facilitate the fundamental understanding of the IEI data characteristics and their capacity to reflect the elderly physical and/or mental status in daily life.

Last but not least, we should execute an in-depth analysis of the methods (features and/or models) for reaching an explainable AI [83]. In particular, current benchmark works were focused on the analysis of all day IEI date, not involving detailed analysis on event happening time throughout the day, e.g., opening/closing the refrigerator late at night, which might be related to subjects' symptoms. Moreover, how to combine classic ML models and DL models to reach a better understanding of the IEI data and building a trustable AI system will be needed along the future path.

VI. Conclusion

In this study, we proposed a novel framework based on ML for analyzing the behavior of elderly in daily life via their usage of smart appliances. Three paradigms of ML were investigated and compared, i.e., functionals, bag-of-behavior-words, and e2e DL. Experimental results demonstrated the feasibility of the proposed systems, which achieved a best UAR of 58.7 % by the BoBW approach for a binary classification task (*Symptom* or *Nonsymptom*, chance level: 50 % UAR).

Acknowledgment

The authors would like to thank the colleagues who contributed to collect the IEI data for this study.

REFERENCES

[1] A. Schmidt and K. Van Laerhoven, "How to build smart appliances?" *IEEE Pers. Commun.*, vol. 8, no. 4, pp. 66–71, Aug. 2001.

[2] S. Nistor, J. Wu, M. Sooriyabandara, and J. Ekanayake, "Capability of smart appliances to provide reserve services," *Appl. Energy*, vol. 138, pp. 590–597, Jan. 2015.

[3] C. B. Kobus, E. A. Klaassen, R. Mugge, and J. P. Schoormans, "A real-life assessment on the effect of smart appliances for shifting households' electricity demand," *Appl. Energy*, vol. 147, pp. 335–343, Jun. 2015.

[4] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "Power demand control scenarios for smart grid applications with finite number of appliances," *Appl. Energy*, vol. 162, pp. 83–98, Jan. 2016.

[5] K. Paridari, A. Parisio, H. Sandberg, and K. H. Johansson, "Robust scheduling of smart appliances in active apartments with user behavior uncertainty," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 247–259, Jan. 2016.

[6] J. Wang, H. Zhang, and Y. Zhou, "Intelligent under frequency and under voltage load shedding method based on the active participation of smart appliances," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 353–361, Jan. 2017.

[7] F. Samie, L. Bauer, and J. Henkel, "From cloud down to things: An overview of machine learning in Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4921–4934, Jun. 2019.

[8] I. Ahmed, A. Ahmad, F. Piccialli, A. K. Sangaiah, and G. Jeon, "A robust features-based person tracker for overhead views in industrial environment," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1598–1605, Jun. 2018.

[9] W. Li, T. Logenthiran, V.-T. Phan, and W. L. Woo, "Implemented IoT-based self-learning home management system (SHMS) for Singapore," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2212–2219, Jun. 2018.

[10] W. Hu, Y. Wen, K. Guan, G. Jin, and K. J. Tseng, "iTCM: Toward learning-based thermal comfort modeling via pervasive sensing for smart buildings," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 4164–4177, Oct. 2018.

[11] W. Zhang, W. Hu, and Y. Wen, "Thermal comfort modeling for smart buildings: A fine-grained deep learning approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2540–2549, Apr. 2019.

[12] W. Peng, W. Gao, and J. Liu, "AI-enabled massive devices multiple access for smart city," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7623–7634, Oct. 2019.

[13] T. Sutjarittham, H. H. Gharakheili, S. S. Kanhere, and V. Sivaraman, "Experiences with IoT and AI in a smart campus for optimizing classroom usage," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7595–7607, Oct. 2019.

[14] W.-L. Chen, Y.-B. Lin, F.-L. Ng, C.-Y. Liu, and Y.-W. Lin, "RiceTalk: Rice blast detection using Internet of Things and artificial intelligence technologies," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1001–1010, Feb. 2020.

[15] World Health Organisation (WHO), *World Report on Ageing and Health*. Geneva, Switzerland: WHO Press, 2015.

[16] D. M. Wilson, B. Errasti-Ibarrondo, and G. Low, "Where are we now in relation to determining the prevalence of ageism in this era of escalating population ageing?" *Ageing Res. Rev.*, vol. 51, pp. 78–84, May 2020.

[17] Statista. (2020). *Japan: Age Distribution From 2008 to 2018*. [Online]. Available: https://www.statista.com/statistics/270087/age-distribution-in-japan/

[18] A. Zhavoronkov, P. Mamoshina, Q. Vanhaelen, M. Scheibye-Knudsen, A. Moskalev, and A. Aliper, "Artificial intelligence for aging and longevity research: Recent advances and perspectives," *Ageing Res. Rev.*, vol. 49, pp. 49–66, Jan. 2019.

[19] O. B. Sezer, E. Dogdu, and A. M. Ozbayoglu, "Context-aware computing, learning, and big data in Internet of Things: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 1–27, Feb. 2018.

[20] S. Gahlot, S. Reddy, and D. Kumar, "Review of smart health monitoring approaches with survey analysis and proposed framework," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2116–2127, Apr. 2019.

[21] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 81–94, Mar. 2016.

[22] H. Mshali, T. Lemlouma, M. Moloney, and D. Magoni, "A survey on health monitoring systems for health smart homes," *Int. J. Ind. Ergon.*, vol. 66, pp. 26–56, Jul. 2018.

[23] V. Nathan *et al.*, "A survey on smart homes for aging in place: Toward solutions to the specific needs of the elderly," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 111–119, Sep. 2018.

[24] S. Deep, X. Zheng, C. Karmakar, D. Yu, L. Hamey, and J. Jin, "A survey on anomalous behavior detection for elderly care using dense-sensing networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 352–370, 1st Quart., 2020.

[25] D. Ravì, C. Wong, B. Lo, and G.-Z. Yang, "A deep learning approach to on-node sensor data analytics for mobile or wearable devices," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 56–64, Jan. 2017.

[26] J. Venkatesh, B. Aksanli, C. S. Chan, A. S. Akyurek, and T. S. Rosing, "Modular and personalized smart health application design in a smart city environment," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 614–623, Apr. 2018.

[27] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. De Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8553–8562, Oct. 2019.

[28] G. Yang *et al.*, "An IoT-enabled stroke rehabilitation system based on smart wearable armband and machine learning," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–10, 2018.

[29] M. Chen, Y. Li, X. Luo, W. Wang, L. Wang, and W. Zhao, "A novel human activity recognition scheme for smart health using multilayer extreme learning machine," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1410–1418, Apr. 2019.

[30] Y. Xiao, Y. Jia, X. Cheng, J. Yu, Z. Liang, and Z. Tian, "I can see your brain: Investigating home-use electroencephalography system security," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6681–6691, Aug. 2019.

[31] G. Paragliola and A. Coronato, "Gait anomaly detection of subjects with Parkinson's disease using a deep time series-based approach," *IEEE Access*, vol. 6, pp. 73280–73292, 2018.

[32] H. Xu, Y. Pan, J. Li, L. Nie, and X. Xu, "Activity recognition method for home-based elderly care service based on random forest and activity similarity," *IEEE Access*, vol. 7, pp. 16217–16225, 2019.

[33] M. Awais, L. Chiari, E. A. F. Ihlen, J. L. Helbostad, and L. Palmerini, "Physical activity classification for elderly people in free-living conditions," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 197–207, Jan. 2019.

[34] M. Gochoo, T.-H. Tan, V. Velusamy, S.-H. Liu, D. Bayanduuren, and S.-C. Huang, "Device-free non-privacy invasive classification of elderly travel patterns in a smart house using PIR sensors and DCNN," *IEEE Sensors J.*, vol. 18, no. 1, pp. 390–400, Jan. 2018.

[35] M. Gochoo, T.-H. Tan, S.-H. Liu, F.-R. Jean, F. S. Alnajjar, and S.-C. Huang, "Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and DCNN," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 693–702, Mar. 2019.

[36] C. Tsirmpas, A. Anastasiou, P. Bountris, and D. Koutsouris, "A new method for profile generation in an Internet of Things environment: An application in ambient-assisted living," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 471–478, Dec. 2015.

[37] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.

[38] M. Roubens, "Fuzzy clustering algorithms and their cluster validity," *Eur. J. Oper. Res.*, vol. 10, no. 3, pp. 294–301, 1982.

[39] Y. Hbali, S. Hbali, L. Ballihi, and M. Sadgal, "Skeleton-based human activity recognition for elderly monitoring systems," *IET Comput. Vis.*, vol. 12, no. 1, pp. 16–26, 2017.

[40] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.

[41] Y. Chen, L. Yu, K. Ota, and M. Dong, "Robust activity recognition for aging society," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1754–1764, Nov. 2018.

[42] N. E. Huang *et al.*, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[43] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 1992.

[44] K. Aoki *et al.*, "Early detection of lower MMSE scores in elderly based on dual-task gait," *IEEE Access*, vol. 7, pp. 40085–40094, 2019.

[45] L. Zhou, G. Wang, C. Jia, and Z. Ma, "Being left-behind, mental disorder, and elderly suicide in rural China: A case–control psychological autopsy study," *Psychol. Med.*, vol. 49, no. 3, pp. 458–464, 2019.

[46] K. Yoshiuchi, Y. Yamamoto, H. Niwamoto, T. Watsuji, H. Kumano, and T. Kuboki, "Behavioral power-law exponents in the usage of electric appliances correlate mood states in the elderly," *Int. J. Sport Health Sci.*, vol. 1, no. 1, pp. 41–47, 2003.

[47] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[48] F. Eyben, *Real-Time Speech and Music Classification by Large Audio Feature Space Extraction*. Cham, Switzerland: Springer Int., 2015.

[49] K. Qian *et al.*, "Teaching machines to know your depressive state: On physical activity in health and major depressive disorder," in *Proc. EMBC*, Berlin, Germany, 2019, pp. 3592–3595.

[50] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Appl. Intell.*, vol. 7, no. 1, pp. 39–55, 1997.

[51] M. N. Murty and V. S. Devi, *Pattern Recognition: An Algorithmic Approach*. Dordrecht, The Netherlands: Springer, 2011.

[52] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, 1936.

[53] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[54] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[55] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[56] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[57] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[58] Z. S. Harris, "Distributional structure," *WORD*, vol. 10, nos. 2–3, pp. 146–162, 1954.

[59] J. Sivic and A. Zisserman, "Efficient visual search of videos cast as text retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 591–606, Apr. 2009.

[60] F. Weninger, P. Staudt, and B. Schuller, "Words that fascinate the listener: Predicting affective ratings of on-line lectures," *Int. J. Distance Educ. Technol.*, vol. 11, no. 2, pp. 110–123, 2013.

[61] S. Rawat, P. F. Schulam, S. Burger, D. Ding, Y. Wang, and F. Metze, "Robust audio-codebooks for large-scale event detection in consumer videos," in *Proc. INTERSPEECH*, Lyon, France, 2013, pp. 2929–2933.

[62] M. Schmitt, F. Ringeval, and B. Schuller, "At the border of acoustics and linguistics: Bag-of-audio-words for the recognition of emotions in speech," in *Proc. INTERSPEECH*, San Francisco, CA, USA, 2016, pp. 495–499.

[63] K. Qian *et al.*, "A bag of wavelet features for snore sound classification," *Ann. Biomed. Eng.*, vol. 47, no. 4, pp. 1000–1011, 2019.

[64] K. Qian *et al.*, "Automatic detection of major depressive disorder via a bag-of-behavior-words approach," in *Proc. ISICDM*, Xi'an, China, 2019, pp. 71–75.

[65] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. ACM-SIAM SODA*, New Orleans, LA, USA, 2007, pp. 1027–1035.

[66] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[67] S. Pancoast and M. Akbacak, "Bag-of-audio-words approach for multimedia event classification," in *Proc. INTERSPEECH*, Portland, OR, USA, 2012, pp. 2105–2108.

[68] S. Dieleman and B. Schrauwen, "End-to-end learning for music audio," in *Proc. ICASSP*, Florence, Italy, 2014, pp. 6964–6968.

[69] G. Trigeorgis *et al.*, "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network," in *Proc. ICASSP*, Shangai, China, 2016, pp. 5200–5204.

[70] M. Schmitt and B. Schuller, "End-to-end audio classification with small datasets—Making it work," in *Proc. EUSIPCO*, 2019, pp. 1–5.

[71] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Proc. NIPS*, Denver, CO, USA, 1989, pp. 396–404.

[72] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.

[73] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[74] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS Deep Learn. Represent. Learn. Workshop*, Montreal, QC, Canada, 2014, pp. 1–9.

[75] S. Hochreiter *et al.*, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Neural Networks*, J. F. Kolen and S. C. Kremer, Ed. Piscataway, NJ, USA: IEEE, 2001, pp. 237–244.

[76] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[77] B. Schuller, S. Steidl, and A. Batliner, "The INTERSPEECH 2009 emotion challenge," in *Proc. INTERSPEECH*, Brighton, U.K., 2009, pp. 312–315.

[78] K. Qian, "Automatic general audio signal classification," Ph.D. dissertation, Lehrstuhl für Mensch-Maschine-Kommunikation Technische Universität München, München, Germany, 2018.

[79] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm

[80] M. Schmitt and B. W. Schuller, "openXBOW—Introducing the PASSAU open-source crossmodal bag-of-words toolkit," *J. Mach. Learn. Res.*, vol. 18, no. 96, pp. 1–5, 2017.

[81] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. NIPS*, Montreal, QC, Canada, 2014, pp. 2672–2680.

[82] J. Kim, T. Nakamura, H. Kikuchi, K. Yoshiuchi, T. Sasaki, and Y. Yamamoto, "Covariation of depressive mood and spontaneous physical activity in major depressive disorder: Toward continuous monitoring of depressive mood," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1347–1355, Jul. 2015.

[83] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

Kun Qian (Senior Member, IEEE) received the doctoral degree for his study on automatic general audio signal classification in electrical engineering and information technology from the Technische Universität München (TUM), Munich, Germany, in 2018.

He is currently working as a JSPS Postdoctoral Research Fellow with the Educational Physiology Laboratory, Graduate School of Education, University of Tokyo, Tokyo, Japan. He was also sponsored by fellowships to conduct cooperative research with the Nanyang Technological University, Singapore, Tokyo Institute of Technology (Tokyo Tech), Tokyo, and the Carnegie Mellon University, Pittsburgh, PA, USA. He (co-)authored more than 50 publications in peer reviewed journals, and conference proceedings having received more than 700 citations (H-index 16). His main research interests include signal processing, machine learning, biomedical engineering, and deep learning.

Dr. Qian serves as an Associate Editor for *Fontiers in Digital Health*, and is the leading organizer of the special session on computer audition for healthcare in the ICASSP 2021, Toronto, Canada. He reviews regularly for many prestigious journals (e.g., IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING). He was also the Reviewer for the past ICASSP/INTERSPEECH/EMBC conferences.

Tomoya Koike (Student Member, IEEE) received the B.Sc. degree from Kobe University, Kobe, Japan, in 2020. He is currently pursuing the master's degree with the Graduate School of Education, University of Tokyo, Tokyo, Japan.

He is the main author of the open source toolkit *deepSELF*, and has successfully published academic papers in a series of prestigious conferences, such as EMBC and INTERSPEECH. His research interests include machine learning, deep learning, and healthcare applications.

Kazuhiro Yoshiuchi received the M.D. degree from the School of Medicine, University of Tokyo, Tokyo, Japan, in 1991, and the Ph.D. degree in medical science from the University of Tokyo in 1998.

He is currently an Associate Professor with the Department of Stress Sciences and Psychosomatic Medicine, Graduate School of Tokyo, University of Tokyo, Tokyo. He is also the Director of the Department of Psychosomatic Medicine, University of Tokyo Hospital. His current research interests include ecological momentary assessment and ecological momentary intervention in stress-related diseases.

Björn W. Schuller (Fellow, IEEE) received the diploma, doctoral, and habilitation degrees in electrical engineering and information technology and the Adjunct Teaching Professorship in the subject area of signal processing and machine intelligence from the Technische Universität München, Munich, Germany, in 1999, 2006, and 2012, respectively.

He is a Tenured Full Professor heading the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany, and a Professor of Artificial Intelligence heading GLAM, Department of Computing, Imperial College London, London, U.K. He (co-)authored five books and more than 900 publications in peer reviewed books, journals, and conference proceedings leading to more than 33 000 citations (H-index 85).

Dr. Schuller is the field Chief Editor of *Frontiers in Digital Health*, the former Editor-in-Chief of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, a President-Emeritus of AAAC, a Fellow of the Golden Core Awardee of the IEEE Computer Society and ISCA, and a Senior Member of ACM.

Yoshiharu Yamamoto (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in education from the University of Tokyo, Tokyo, Japan, in 1984, 1986, and 1990, respectively.

Since 2000, he has been a Professor with the Graduate School of Education, University of Tokyo, where he is teaching and researching physiological bases of health sciences and education. He (co-)authored more than 230 publications in peer reviewed books, journals, and conference proceedings leading to more than 11 000 citations (H-index 55). His research interests include biomedical signal processing, nonlinear and statistical biodynamics, and health informatics.

Dr. Yamamoto is currently an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, and an Editorial Board Member of the Technology and *Biomedical Physics and Engineering Express*. He is also the President of the Healthcare IoT Consortium, Japan.